Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia

van Dijk, Lisanne V.; Thor, Maria; Steenbakkers, Roel J. H. M.; Apte, Aditya; Zhai, Tian-Tian; Borra, Ronald; Noordzij, Walter; Estilo, Cherry; Lee, Nancy; Langendijk, Johannes A.

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2018.06.012

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Radiotherapy and Oncology 128 (2018) 459–466

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Head and neck cancer

Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia

Lisanne V. van Dijk [a,*], Maria Thor [b], Roel J.H.M. Steenbakkers [a], Aditya Apte [b], Tian-Tian Zhai [a], Ronald Borra [c], Walter Noordzij [d], Cherry Estilo [e], Nancy Lee [f], Johannes A. Langendijk [a], Joseph O. Deasy [b], Nanna M. Sijtsema [a]

[a] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, The Netherlands; [b] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, United States; [c] Department of Radiology, University of Groningen, University Medical Center Groningen; [d] Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, The Netherlands; [e] Department of Surgery, Memorial Sloan Kettering Cancer Center; and [f] Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, United States

ARTICLE INFO

Article history:
Received 28 March 2018
Received in revised form 4 June 2018
Accepted 6 June 2018
Available online 26 June 2018

Keywords:
Xerostomia
NTCP
Image biomarkers
Head and neck cancer
Magnetic Resonance Imaging
Radiomics

ABSTRACT

Purpose: This study investigated whether Magnetic Resonance image biomarkers (MR-IBMs) were associated with xerostomia 12 months after radiotherapy (Xer_{12m}) and to test the hypothesis that the ratio of fat-to-functional parotid tissue is related to Xer_{12m}. Additionally, improvement of the reference Xer_{12m} model based on parotid gland dose and baseline xerostomia, with MR-IBMs was explored.

Methods: Parotid gland MR-IBMs of 68 head and neck cancer patients were extracted from pre-treatment T1-weighted MR images, which were normalized to fat tissue, quantifying 21 intensity and 43 texture image characteristics. The performance of the resulting multivariable logistic regression models after bootstrapped forward selection was compared with that of the logistic regression reference model. Validity was tested in a small external cohort of 25 head and neck cancer patients.

Results: High intensity MR-IBM P90 (the 90th intensity percentile) values were significantly associated with a higher risk of Xer_{12m}. High P90 values were related to high fat concentration in the parotid glands. The MR-IBM P90 significantly improved model performance in predicting Xer_{12m} (likelihood-ratio-test; $p = 0.002$), with an increase in internally validated AUC from 0.78 (reference model) to 0.83 (P90). The MR-IBM P90 model also outperformed the reference model (AUC = 0.65) on the external validation cohort (AUC = 0.83).

Conclusion: Pre-treatment MR-IBMs were associated to radiation-induced xerostomia, which supported the hypothesis that the amount of predisposed fat within the parotid glands is associated with Xer_{12m}. In addition, xerostomia prediction was improved with MR-IBMs compared to the reference model.

Xerostomia is one of the most frequently reported side-effects following radiotherapy for head and neck cancer, and has a major impact on quality of life [1,2]. Normal Tissue Complication Probability (NTCP) models have been developed to predict radiation-induced xerostomia and have demonstrated a clear relationship with parotid gland dose and baseline patient-rated xerostomia [3,4]. Nevertheless, substantial unexplained variance in predicting xerostomia remains. Better understanding of the aetiology of radiation-induced xerostomia is necessary to advance towards more individualised treatments and better sparing of normal tissues by further dose optimization, by means of new radiation techniques, such as proton therapy [5,6] and Magnetic Resonance Imaging (MRI) guided radiation [7].

Tumour-based image biomarkers (IBMs), which are shape, intensity and texture characteristics extracted from images, can contribute to the prediction of overall, disease-free and progression-free survival [8–13]. However, the role of these IBMs in normal tissues to predict radiation-induced toxicities is less explored, while these are imperative in supporting treatment decisions [5].

Our previous study based on IBMs from pre-treatment CT images, demonstrated that high heterogeneous parotid gland tissue, was associated with a higher probability of developing late xerostomia [14]. Qualitative evaluation of the parotid glands suggested that the predictive CT-IBM indicated the ratio between fatty and functional parotid parenchyma tissue. In a subsequent study, we showed that patients with low metabolic parotid glands,

* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands. Fax: +31 503613672.
E-mail address: l.v.van.dijk@umcg.nl (L.V. van Dijk).

https://doi.org/10.1016/j.radonc.2018.06.012
0167-8140/

quantified in pre-treatment ^{18}FDG-PET IBMs, were more likely to develop late xerostomia. These associations also suggested that the non-functional (which can be fatty tissue) to functional tissue ratio is an important pre-treatment characteristic to improve prediction of xerostomia [15].

MRI is superior in imaging soft tissue contrast and therefore more accurate in differentiating fat from the parenchymal gland tissue compared to CT and ^{18}FDG-PET [16]. Hence, investigating the pre-treatment MR-IBMs of the parotid glands could, therefore, potentially provide better information for predicting late xerostomia.

The purpose of this study was to test whether MR-IBMs extracted from T1-weighted MRI scans were associated with the development of xerostomia 12 months after radiotherapy (Xer_{12m}) and to investigate whether MR-IBMs can improve the xerostomia prediction model based on parotid gland dose and baseline xerostomia only. The predictive MR-IBMs were evaluated to test the hypothesis that the fat-to-functional parenchymal parotid tissue ratio is related to Xer_{12m}. The findings were externally validated in an independent cohort.

Materials and methods

Patient demographics and treatment

The training and test cohort included head and neck cancer patients that were treated with definitive radiotherapy with or without concurrent chemotherapy or cetuximab between September 2012 and December 2014 at the University Medical Center Groningen (UMCG), and between October 2010 and March 2016 at Memorial Sloan Kettering Cancer Center (MSKCC), respectively. All patients were treated with Intensity-Modulated Radiation Therapy (IMRT) or Volumetric Arc Therapy (VMAT) using a simultaneous integrated boost (SIB) technique. The parotid glands were spared as much as possible. Patients received a total therapeutic dose of 70 Gy over 6–7 weeks. Most patients received bilateral neck radiation with a prophylactic dose of 54.25 Gy. Details about the radiotherapy regimens used are described in detail in previous studies [14,17].

Patients were excluded if they had salivary gland tumours or underwent surgery or radiotherapy in the head and neck area prior to or within one year after treatment. Moreover, patients without late follow-up data were excluded. Furthermore, MRI scan quality was visually evaluated, and if scans had considerable noise, limiting both visualisation of the parotid glands and reliable estimation of the local image intensity, patients were excluded. The final number of patients was 68 and 25 in the UMCG and MSKCC cohorts, respectively.

Endpoints

The primary endpoint was patient-rated moderate-to-severe late xerostomia (Xer_{12m}). In the UMCG cohort, this corresponds to the 2 highest scores of the 4-point Likert scale of the EORTC QLQ-H&N35 questionnaire and was consistently scored 12 months after treatment, which is part of a Standard Follow-up Program (SFP) for Head and Neck Cancer Patients (NCT02435576), as described in previous studies [4,18].

In the MSKCC cohort, xerostomia was scored with multiple questions with a 0–10 scale [19,20] (see Supplemental Materials 1). Xerostomia scores were collected between 6 and 17 months after treatment (mean ± SD: 11.0 ± 2.5 months). Moderate-to-severe xerostomia was considered if any of the questions was scored 6 or higher.

MRI acquisition and standardisation

In the UMCG, MR images were acquired in treatment position on a single scanner (MAGNETOM Aera 1.5 T scanner, Siemens Medical Systems, Knoxville, TN, USA) approximately 2 weeks before the start of radiotherapy (Spine 32, flexible 4 and 18 channel coils) for delineation purposes. T1-weighted Turbo Spin Echo (TSE) images (TE: 22 ms; TR: 457–606 ms) were acquired for all patients with a resolution of 0.36 × 0.36 × 4.00 mm without the use of intravenous contrast agents or fat-suppression.

In MSKCC, pre-treatment MR-images were acquired on MRI scanners of different manufacturers (GE, Phillips, Siemens) and scanners with field strength of 1.5 T (13 patients) and 3 T (12 patients). The resolution of the non-contrast enhanced T1-weighted TSE images (TE: 8–20 ms; TR: 400–697 ms) ranged from 0.35x0.35 to 1.01 x1.01 mm in-plane and the slice thickness from 3.0 to 5.0 mm.

The MRI intensity values of similar tissue types vary between scans. Therefore, only relative intensities within one scan can be compared. To make a comparison of the relative intensities between patients possible, scans had to be standardised. In this study, fat T1 characteristics were assumed consistent between patients, and should, consequently, have similar MR-intensity values. Subcutaneous fat was delineated in both the right and left cheek area in a minimum of 4 slices at the level of the parotid glands of all patients (Fig. 1I). The fat area was delineated laterally of the parotid gland, the masseter muscle and lip muscles, where the area is delineated as large as possible while excluding non-fat related structures. Subsequently, the MR images were multiplied by a fixed value, which was arbitrarily chosen to 350, and divided by the average subcutaneous fat intensity value. This

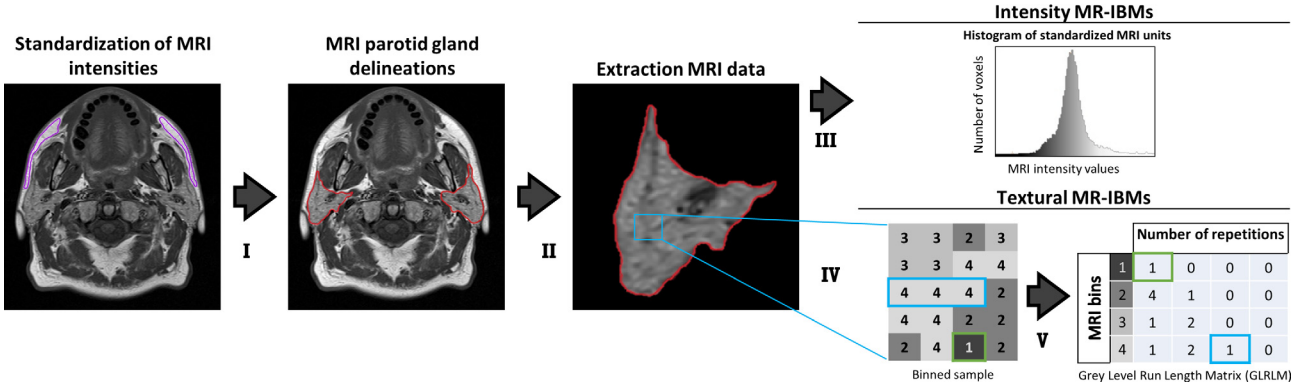

Fig. 1. MR-IBM extraction process. (I) MR scans were standardised with the average MR intensity, obtained from two delineated fat regions (purple). (II) MRI delineated parotid glands were extracted. (III) Intensity MR-IBMs were directly extracted. (IV) The MR intensities were binned. (V) Textural MR-IBMs were extracted from matrices reflection grey level transitions or repetitions.

approach is a simplified tissue-based MRI Intensity standardisation [21]; to our knowledge, no MRI standardisation approaches are known for the head and neck area or radiomics purposes.

Candidate MRI-IBMs, dose and clinical parameters

Parotid glands were delineated for clinical planning purposes on the planning CT, according to guidelines of Brouwer et al. [22]. Dosimetric parameters were extracted from these volumes. MR images were rigidly matched to the CT scans, and the CT contours were transferred to the MR images. The MRI parotid delineations were manually corrected where necessary in both datasets.

MRI characteristics of the delineated parotid glands were quantified in intensity and texture MR-image biomarkers (MR-IBMs). Intensity MR-IBMs represent first-order MR-intensity characteristics, such as the mean, minimum, maximum, standard deviation and root mean square of the MR-intensity values.

Furthermore, the MR-intensity heterogeneity was quantified by the textural MR-IBMs. These were extracted from the grey level co-occurrence matrix (GLCM) [23] and grey level run-length matrix (GLRLM) [24,25], where GLCM describes the grey level transitions, GLRLM describes the directional grey level repetitions. The texture IBMs were evaluated in 2D only, which means that the average of MR-IBM values from GLCM and GLRLM of 4 independent directions in-plain were used. Intensity values were discretized from 0 to 450 with a bin size of 25 standardised MR-units [26].

For the complete list of the 21 intensity and 43 textural MR-IBMs (25 GLCM and 18 GLRLM) see Supplementary data 2. The extraction process (Fig. 1) was performed in MATLAB 2014a and all definitions and formulas were according to the 'Image biomarker standardisation initiative' [27].

Multivariable analysis and model performance

Reference model

A multivariable logistic regression reference model based on the mean dose to the both parotid glands and patient-reported xerostomia at the start of radiotherapy ($Xer_{baseline}$) was fitted in the training cohort [3,4]. $Xer_{baseline}$ was dichotomized as none vs. any in the UMCG dataset and larger than 1 in the MSKCC dataset.

Intensity and textural MR-IBMs selection

To understand the contribution of the different types of MR-IBMs to the reference model, intensity and textural based MR-IBM models were considered separately. Model training was performed in the UMCG cohort only. MR-IBM values were normalized by subtracting each value by the average IBM value and then dividing by standard deviation of that IBM variable.

A pre-selection based on (Pearson) correlation was performed to reduce the effects of overfitting and multicollinearity. If the correlation between two candidate MR-IBMs was larger than 0.80, only the variable with the highest association with Xer_{12m} was selected.

Multivariable logistic analysis of the pre-selected MR-IBMs was performed together with the mean parotid dose and $Xer_{baseline}$. Based on largest significant log-likelihood differences, step-wise forward selection was used to select MR-IBM predictors [28] (p-value <0.01).

The internal validity of the variable selection was estimated by repeating the entire variable selection procedure (variable normalization, pre-selection and forward selection) 1000 times with a bootstrap procedure with replacement (i.e. with repetition and same sample size). The most frequently selected variables were considered the final model. Model optimism was estimated by calculating the average difference between the performance of the models in each bootstrap and in the original sample, as suggested by the TRIPOD statement [29].

Trained on the UMCG cohort, the MR-IBM models were externally validated in the MSKCC cohort. The model performance measures were the area under the ROC (receiver operating characteristic) curve (AUC), Nagelkerke's R^2 and the discrimination slope. Model calibration was tested with the average slope and intercept of the models trained on the bootstrap samples that were tested on the original data. The coefficients were corrected for optimism accordingly. In addition, the model improvement was determined with the Likelihood-ratio test, Integrated Discrimination Improvement (IDI) and DeLong's test (testing if AUC significantly improves). The R-package Regression Modelling Strategies (version 4.3-1) [30] and pROC (version 1.8) were used for these purposes.

Inter-variable relationships

The relation between predictive IBMs and $Xer_{baseline}$ were investigated with the Pearson correlation and univariable logistic analysis, respectively.

Results

Patient characteristics are depicted in Table 1. Generally, all patients received bilateral irradiation. The majority of the patients had oropharyngeal carcinomas and did not report any $Xer_{baseline}$ (59% in the UMCG cohort; 56% in the MSKCC cohort). Moderate-to-severe xerostomia 12 months after radiotherapy (Xer_{12m}) was reported by 34 (50%) of the 68 patients in the UMCG cohort and by 10 (40%) of 25 patients from the MSKCC cohort. In addition, the average (±standard deviation) mean PG dose was 31.8 ± 10.9 Gy and 22.0 ± 8.8 Gy in the UMCG and MSKCC cohort, respectively. Mean dose to both parotid glands performed slightly better than the contra-lateral gland in this cohort, probably due to the tumour location (oropharynx) and advance N-stage, resulting in comparable contra- and ipsi-lateral doses.

The reference model based on mean PG dose and $Xer_{baseline}$ was fitted to the training dataset. The model characteristics and the performance measures (AUC = 0.81 (95%CI:0.71–0.91), R^2 = 0.39) are depicted in Tables 2 and 3. The reference model showed reduced performance in the external dataset ($AUC_{external.val.}$ = 0.65 (0.41–0.88), $R^2_{external.val.}$ = 0.07).

Resulting from the bootstrapped variable selection of the intensity MR-IBMs, the 90th intensity percentile (P90) of standardised MRI-units to fat tissue was most frequently selected (175 times of 1000 bootstrapped samples; see Supplementary data 3 for frequency plots). This MR-IBM had both a univariable (OR = 1.03 (1.01–1.05); p = 0.004) and multivariable (Table 2) association with Xer_{12m}. The positive regression coefficient reveals that high P90 is associated with higher risk of developing Xer_{12m} (Table 2). Fig. 2 depicts example patients with high and low P90 values.

The P90 added significantly to the variables of the reference model (Likelihood-ratio test; p = 0.002; IDI; p = 0.004), and resulted in a substantial and significant improvement of the model performance measures (DeLong's test; p = 0.04), increasing the reference model's AUC of 0.81 (95%CI:0.71–0.91; R^2 = 0.39, $AUC_{internal.val.}$ = 0.78) to 0.88 (0.79–0.96; R^2 = 0.51, $AUC_{internal.val.}$ = 0.84) for the intensity MR-IBM model (Table 3). The NTCP-curves for different P90 values are depicted in Fig. 3.

The performance of the P90 model remained good when externally validated in the MSKCC dataset ($AUC_{external.val.}$ = 0.83 (0.66–0.99), $R^2_{external.val.}$ = 0.36). In addition, univariable analysis in the external dataset showed a significant association of P90 with Xer_{12m} (p = 0.039).

From the texture MR-IBMs, the Grey Level Non-uniformity Normalized (GLN_{nor}) was most often selected (91 of 1000 bootstrapped

Table 1
Patient characteristics.

Characteristics	UMCG		MSKCC	
	N = 68	%	*N* = 25	%
Sex				
Female	27	40	5	20
Male	41	60	20	80
Age				
18–65 years	47	69	23	92
>65 years	21	31	2	8
Tumour site				
Oropharynx	42	62	17	68
Nasopharynx	5	7	7	28
Hypopharynx	6	9	–	–
Larynx	10	15	–	–
Oral cavity	2	3	–	–
Other	3	4	1	4
Tumour classification				
Tx			2	8
T1	11	16	8	32
T2	20	29	8	32
T3	16	24	4	16
T4	21	31	3	12
Node classification				
N0	19	28	7	28
N1	6	9	4	16
N2	42	62	14	56
N3	1	1	–	–
Systemic treatment				
Yes	42	62	22	88
No	22	32	3	12
Cetuximab	4	6	–	–
Treatment technique				
IMRT	60	88	15	60
VMAT	8	12	10	40
Neck irradiation				
Bilateral	62	91	20	80
Unilateral	1	1	4	16
No	5	7	1	4
Baseline Xerostomia				
No	40	59	14	56
Any	28	41	11	44

Abbreviations: IMRT: Intensity-Modulated Radiation Therapy; VMAT: Volumetric Arc Therapy; UMCG: University Medical Center Groningen; MSKCC: Memorial Sloan Kettering Cancer Center.

samples; see Supplementary data 3). Derived from the GLRLM, this texture MR-IBM is high when high concentrations of runs with the same grey level are present in the volume of interest (for formula see Supplementary data 2). This texture MR-IBM had both a negatively significant univariable (OR = 0.34, 95%CI 0.20–0.74; $p = 0.004$) and multivariable (Table 2) association with Xer_{12m}, indicating that low GLN_{nor} values were related with a higher risk of xerostomia.

When adding GLN_{nor} to the reference model, model performance significantly improved (Likelihood-ratio test; $p = 0.002$; IDI; $p < 0.001$). The performance of the resulting texture MR-IBM model (Table 2) was good (AUC: 0.88 (0.79–0.96), $R^2 = 0.52$; $AUC_{internal.val.} = 0.84$) and significantly improved compared to the reference model (DeLong's test; $p = 0.03$). The NTCP curves are depicted in Fig. 3.

On external validation, the texture MR-IBM model performed well ($AUC_{external.val.} = 0.83(0.67–0.99)$, $R^2_{external.val.} = 0.31$). Univariable analysis also showed a significant association of GLN_{nor} with Xer_{12m} ($p = 0.036$).

The internal validation calibration slope and intercept showed reasonable goodness-of-fit for both the intensity and texture models (Table 3). In addition, all models were fitted to the combined

Table 2
Model characteristics of prediction models with and without MR-IBMs trained in UMCG cohort.

	Reference model				Intensity MR-IBM model				Texture MR-IBM model			
	β		OR (95% CI)	*p*-value	*β*		OR (95% CI)	*p*-value	*β*		OR (95% CI)	*p*-value
	Apparent	Corrected			Apparent	Corrected			Apparent	Corrected		
Intercept	−3.311	−2.859			−11.30	−8.540			−3.642	−2.673		
$Xer_{baseline}$	2.351	2.035	10.50 (3.06–36.06)	<0.001	2.734	2.063	15.39 (3.56–66.47)	<0.001	2.736	2.009	15.42 (3.44–69.07)	<0.001
PG dose	0.074	0.064	1.08 (1.01–1.14)	0.015	0.072	0.054	1.07 (1.00–1.15)	0.035	0.080	0.059	1.08 (1.01–1.16)	0.021
P90	–	–	–	–	0.034	0.026	1.03 (1.01–1.06)	0.006	–			
GLN_{nor}	–	–	–	–	–	–	–	–	−1.146	−0.841	0.32 (0.14–0.71)	0.006

Abbreviations: corrected: corrected for optimism with bootstrapping; MR-IBM: Magnetic Resonance Image Biomarker; $Xer_{baseline}$: xerostomia at baseline; PG dose: mean dose to parotid glands; P90: 90th percentile of MR intensities; GLN_{nor}: Grey Level Non-uniformity Normalized; β: regression coefficients; OR: Odds Ratio; CI: Confidence Interval.

Table 3
Performance of prediction models with and without MR-IBMs in training (UMCG) and external validation cohort (MSKCC).

	Reference model			Intensity MR-IBM model			Texture MR-IBM model		
	$Xer_{baseline}$ + PG dose			$Xer_{baseline}$ + PG dose + P90			$Xer_{baseline}$ + PG dose + GLN_{nor}		
	UMCG		MSKCC	UMCG		MSKCC	UMCG		MSKCC
	Apparent	Internal val.	External val.	Apparent	Internal val.	External val.	Apparent	Internal val.	External val.
Area Under the Curve (AUC)	0.81 (0.71–0.91)	0.78	0.65 (0.41–0.88)	0.88 (0.79–0.96)	0.83	0.83 (0.66–0.99)	0.88 (0.79–0.96)	0.84	0.83 (0.67–0.99)
Nagelkerke's R^2	0.39 (0.21–0.39)	0.32	0.07	0.51 (0.41–0.61)	0.39	0.36	0.52 (0.42–0.61)	0.39	0.31
Discrimination slope	0.31	0.27	0.11	0.42	0.35	0.27	0.43	0.37	0.24
Calibration slope (intercept)	–	0.87 (0.00)	0.45 (−0.18)	–	0.75 (0.01)	1.11 (−0.93)	–	0.73 (0.00)	0.94 (−0.28)
Likelihood-ratio test	–	–	–	9.34 (p = 0.002)	–	6.45 (p = 0.011)	9.97 (p = 0.002)	–	5.05 (p = 0.025)
Integrated Discrimination Improvement(IDI)	–	–	–	0.11 (p = 0.004)	–	0.16 (p = 0.035)	0.12 (p < 0.001)	–	0.14 (p = 0.025)
DeLong's test	–	–	–	−1.72 (p = 0.043)	–	−1.87 (0 = 0.030)	−1.87 (p = 0.030)	–	−2.09 (p = 0.018)

Abbreviations: MR-IBM: Magnetic Resonance Image Biomarker; $Xer_{baseline}$: xerostomia at baseline; PG dose: mean dose to parotid glands; P90: 90th percentile of MR intensities; GLN_{nor}: Grey Level Non-uniformity Normalised; UMCG: University Medical Center Groningen; MSKCC: Memorial Sloan Kettering Cancer Center; Internal val.: corrected for optimism with bootstrapping; External val.: externally validated.

dataset (MSKCC + UMCG), and showed similar coefficients and performance measures (Supplementary data 4).

The intensity and texture MR-IBM P90 and GLN_{nor} were highly correlated (r = −0.85 (95%CI: −0.90–−0.78); p < 0.001). In a multivariable analysis, the addition of GLN_{nor} did not add significant information in predicting Xer_{12m} with P90 and vice versa (Likelihood-ratio test; p > 0.27). Furthermore, univariable logistic analysis showed no significant association between $Xer_{baseline}$ and P90 (p = 0.45) or GLN_{nor} (p = 0.29).

Discussion

In previous studies we showed that more heterogeneous CT intensity characteristics and low metabolic ^{18}FDG-PET activity of the parotid glands were related to a higher risk of xerostomia 12 months after radiotherapy (Xer_{12m}) [14,15]. These findings led to the hypothesis that the fat-to-functional parenchymal parotid tissue ratio is an important pre-treatment marker to improve prediction of Xer_{12m}. The results of the current study also support this hypothesis.

Other recent studies also showed that pre-treatment information extracted from CT images, quantifying the parotid gland texture [31] and shape [32], were associated with observed radiation-induced xerostomia. Additionally, several studies have shown associations between xerostomia and parotid gland changes in CT data [33,34]. The current study is novel by investigating pre-treatment MR intensities of the parotid glands, providing high contrast soft-tissue information, in relation to late patient-rated xerostomia.

MRI characteristics of the parotid glands, quantified in pre-treatment MR-IBMs, were significantly associated to Xer_{12m}. Moreover, the Xer_{12m} prediction improved with the addition of MR-IBMs to the reference model using mean parotid glands dose and baseline complaints only (from an AUC of 0.81 to 0.88). These results were also valid in an independent external cohort, where the performance of the reference model ($AUC_{external.val.}$ = 0.65) was low compared to the MR-IBM models ($AUC_{external.val.}$ = 0.83). This underlines the importance of tissue-specific characteristics in predicting and understanding the development of radiation-induced toxicities, which is becoming increasingly important in the selection of patients for more advanced radiation techniques [6,7] and to tailor the treatment to the patient specifically [5].

The most frequently selected intensity MR-IBM was the P90, indicating the 90th percentile of the MR-intensities of the parotid glands. Since the MR-intensity values were standardised to fat, fat tissue can be assumed to have comparable MR-intensity between patients. Since fat has a short T1 relaxation time compared to parenchymal or muscle tissue, it is presented with a high signal intensity in T1-weighted images [35]. Hence high P90 values relate to high fat concentration in the parotid gland. More specifically, if at least 10% of the volume of the parotid gland has high intensity values, patients were at higher risk of developing late xerostomia. However, this volume percentage should be evaluated with caution, since using the simpler 'mean standardized T1 intensity' also significantly improved the reference model (AUC = 0.86; Likelihood-ratio test: p = 0.005). Moreover, 16 of the 21 intensity MR-IBMs and 34 of the 43 texture MR-IBMs also contributed significantly, as single variables, to the reference model in predicting Xer_{12m} (Supplementary data 5). This indicates that other MR-IBMs that are also related to parotid gland intensity and texture, can give similar results as P90 and GLN_{nor}.

The values of the selected texture MR-IBM, GLN_{nor} are low if grey values are equally distributed over all grey levels, i.e. more heterogeneity. Lower GLN_{nor} values were associated with a higher risk of developing xerostomia. This was also demonstrated in a previous study based on CT parotid gland IBMs [14]. Noteworthy,

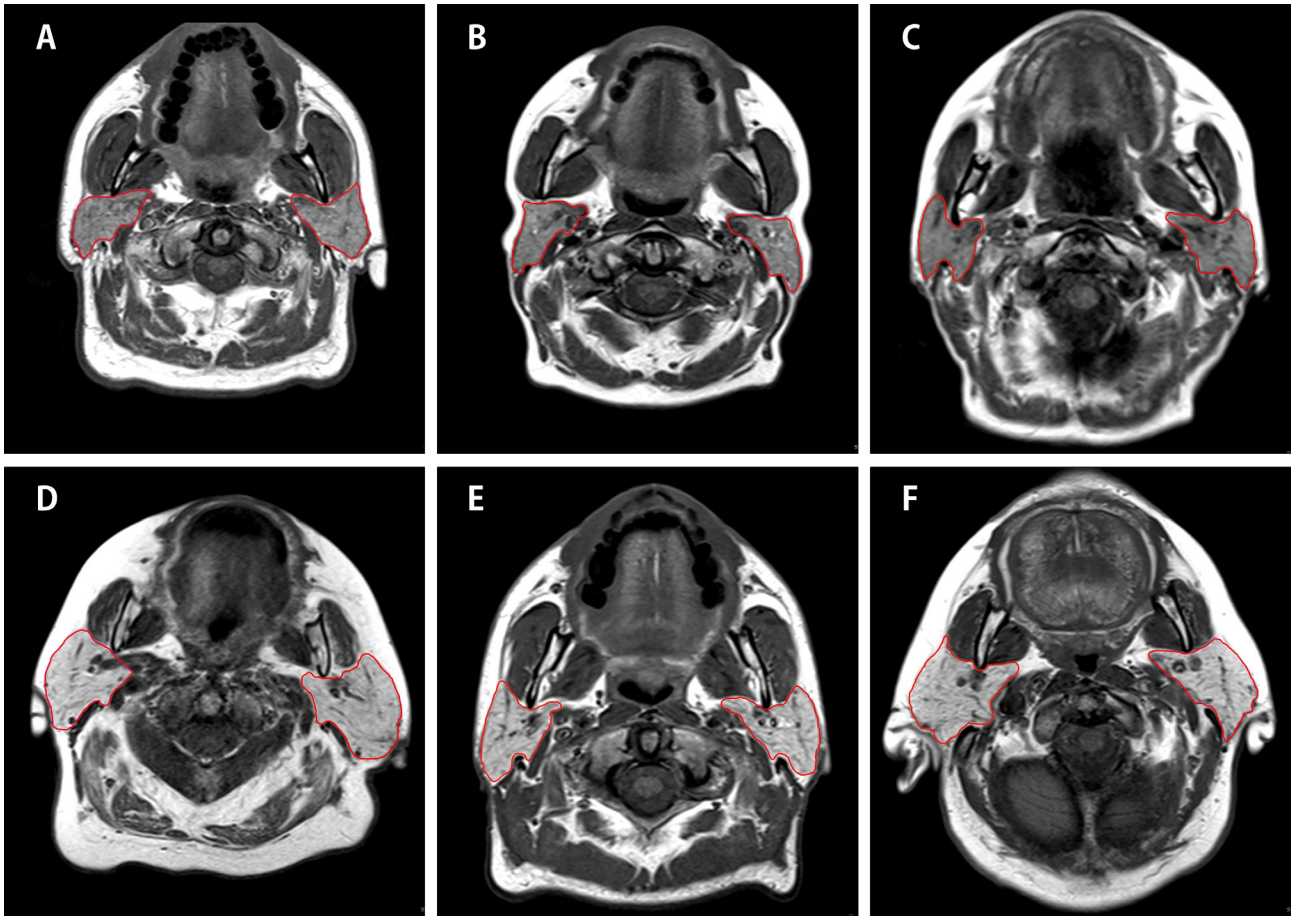

Fig. 2. Examples of patients with low (A–C) and high (D–F) P90 values of the parotid glands. Accordingly, the patients in the bottom row are more at risk of developing late xerostomia than those in the top row.

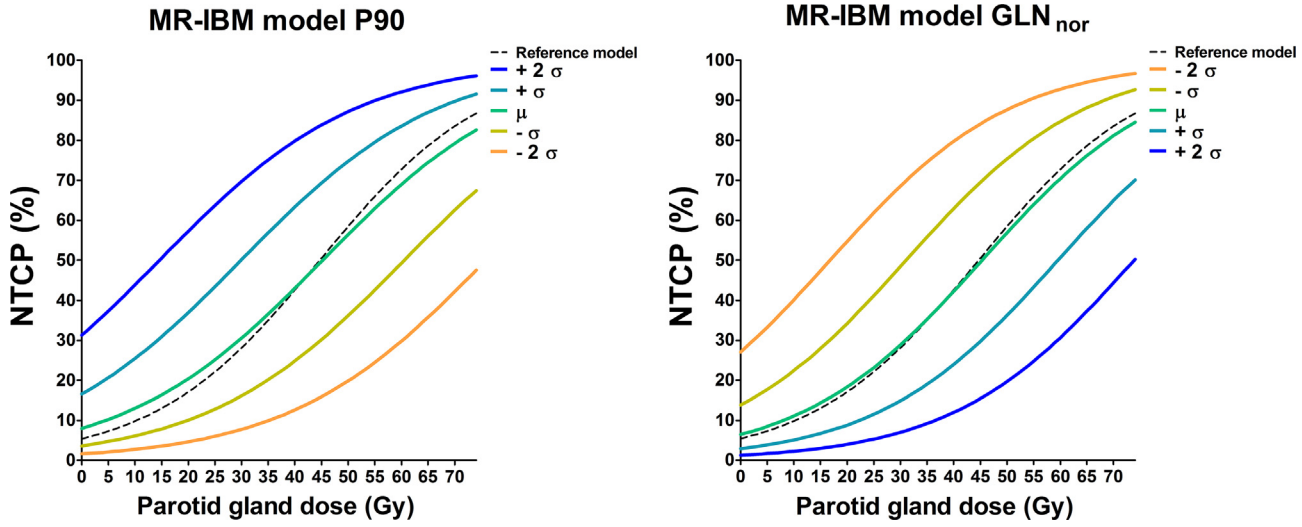

Fig. 3. Normal Tissue Complication Probability (NTCP) models for Xer_{12m} based on the MR-IBM P90 (left) and GLN_{nor} (right) (Table 2). Plotted against the mean dose of both parotid gland, NTCP curves are given for the mean (in green) plus/minus one (turquois/yellow) and two (blue/orange) standard deviations of the P90 ($\mu = 234.89$, $\sigma = 31.80$) and GLN_{nor} ($\mu = 0.18$, $\sigma = 0.03$) values for $Xer_{baseline} = 0$.

the reverse of GLN_{nor} is 'entropy', which was the second most frequently selected intensity MR-IBM (Supplementary data 3). Moreover, GLN_{nor} was highly correlated with P90, which suggests that parotid glands with high (fat related) MR signal intensities were more heterogeneous. More research in larger datasets is necessary to determine whether both characteristics are relevant in the development of xerostomia and the generalisability of P90 and GLN_{nor}, or whether they reflect similar biological information.

The theoretical and qualitative evaluation of the predictive MR-IBMs suggested a relation with the fat concentration and heterogeneity of the parotid gland. In previous work, unrelated to the oncology field, Izumi et al. [36] presented an MRI-based grading of the severity of parotid impairment for patients with Sjögren's syndrome that was based on similar image characteristics: high T1-weigted signal intensity areas (e.g. fat tissue) and heterogeneity in the parotid glands. Another study by Izumi et al. [37] also showed a relationship between increased signal intensities on T1-weighted MR images and impaired parotid function for patients suffering from hyperlipidaemia. The findings of the current and studies suggest that increased fat concentration in the parotid gland, which may be caused by parenchymal changes due to lipid infiltration, can increase the probability of developing xerostomia after radiotherapy.

MRI offers the advantage of non-invasively acquired images with high soft tissue contrast without the use of radiation with respect to CT and PET imaging. However, it is a complex image modality due to the large range of possible acquisition settings, and requires intensity standardisation. This study and previous studies have both indicated that PET and MR IBMs seem to perform better than the CT IBMs in identifying patients that develop xerostomia [14,15]. Studies including IBMs from all three image modalities are necessary to determine which modality is most optimal in this context, or whether they can add to each other in predicting late xerostomia. The analyses of the current study were based on relatively simple T1-weigthed TSE, which is widely used and requires no administration of intravenous contrast agents. However, more sophisticated MRI sequences may better differentiate between fat and functional parotid tissue (i.e. combinations of non- and fat saturated images or functional information (e.g. DIXON, Diffusion Weighted or Dynamic contrast-enhanced imaging)). In addition, IBMs extracted from wavelet transformed images might improve the performance of the models presented in the current study.

Limitations of the present study are the small cohort sizes and the large variability in MR acquisition parameters in the MSKCC cohort compared to the training cohort. Firstly, the resolution had a relative large range in these scans. This can impact the texture IBMs, which depend on the spatial intensity distribution [38,39]. Secondly, patients in the MSKCC cohort were scanned without a thermoplastic mask, resulting in parotid glands deformation due to the music headphones that patients wore during acquisition. Finally, MSKCC scans were acquired with different field strengths, and scanners from different vendors. Even though part of this variability should be captured by MRI standardisation, this can influence the MRI intensity and contrast. An additional limitation is the lack of one-to-one correspondence in xerostomia assessments between the two cohorts. However, a careful matching was performed such that the two moderate-to-severe assessments would be as similar as possible. Despite these limitations, the performance of the MR-IBM models was good when tested in the MSKCC dataset, suggesting that these IBMs were robust to variability in image acquisition parameters. The simplicity of the P90 metric likely contributed to successful validation.

Driven by the hypothesis, the MRI intensity standardisation was linearly performed to ensure similar fat tissue intensities between patients. However, this is in reality not a linear problem [40]. Our approach is simple, and could be regarded as a starting point to improve the standardisation so that not only subcutaneous fat is generalised between patients, but also other tissues, such as muscle. Additionally, mainly due to the presence of field inhomogeneity's, scans can have intensity variations within the scan, for which sophisticated bias field correction algorithms have been developed for brain MR images [41]. The above described corrections were explored for this dataset, however, the effect of these corrections on IBM analysis is currently unknown, and needs further investigation.

In conclusion, the results of the current study support the hypothesis that a high fat concentration, quantified in MR-IBMs, within the parotid glands is related to a higher risk of developing xerostomia 12 months after radiotherapy (Xer_{12m}). The prediction performance of Xer_{12m} based on parotid dose and baseline xerostomia only was improved by the addition of the predictive intensity MR-IBM P90. These results were maintained in a small external validation cohort. MR-IBMs appear to be good candidates to predict the patient-specific response of healthy tissue to radiation dose. However, more research in larger patient cohorts is needed to further validate our conclusions.

Conflict of interest

The authors state that the research presented in this manuscript is free of conflicts of interest.

Acknowledgements

This research was partially funded by the MSK Cancer Center Support Grant/CORE Grant (P30 CA008748).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2018.06.012.

References

[1] Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. Lancet Oncol 2010;11:85–91.
[2] Hawkins PG, Lee JY, Mao Y, Li P, Green M, Worden FP, et al. Sparing all salivary glands with IMRT for head and neck cancer: longitudinal study of patient-reported xerostomia and head-and-neck quality of life. Radiother Oncol 2017;126:68–74.
[3] Houweling AC, Philippens MEP, Dijkema T, Roesink JM, Terhaard CHJ, Schilstra C, et al. A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. Int J Radiat Oncol Biol Phys 2010;76:1259–65.
[4] Beetz I, Schilstra C, Van Der Schaaf A, Van Den Heuvel ER, Doornaert P, Van Luijk P, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:101–6.
[5] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.
[6] Lomax A. Intensity modulation methods for proton radiotherapy. Phys Med Biol 1999;44:185–205.
[7] Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. Radiother Oncol 2008;86:25–9.
[8] Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by (18)F-FDG PET/CT in patients with head and neck squamous cell carcinoma. Eur J Nucl Med Mol Imaging 2014;41:659–67.
[9] Jeong J, Setton JS, Lee NY, Oh JH, Deasy JO. Estimate of the impact of FDG-avidity on the dose required for head and neck radiotherapy local control. Radiother Oncol 2014;111:340–7.
[10] Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. AJR Am J Roentgenol 2014;202:851–8.
[11] Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM. Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. AJR Am J Roentgenol 2014;203:897–903.
[12] Zhai T-T, van Dijk LV, Huang B-T, Lin Z-X, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. Radiother Oncol 2017:256–62.
[13] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5.

[14] van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol 2017;122:185–91.
[15] van Dijk LV, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JGM, Langendijk JA, et al. 18F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. Radiother Oncol 2017;126:89–95.
[16] Burke CJ, Thomas RH, Howlett D. Imaging the major salivary glands. Br J Oral Maxillofac Surg 2011;49:261–9.
[17] Van Der Laan HP, Gawryszuk A, Christianen MEMC, Steenbakkers RJHM, Korevaar EW, Chouvalova O, et al. Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: Treatment planning optimization and clinical introduction. Radiother Oncol 2013;107:282–7.
[18] Vergeer MR, Doornaert PAH, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. Int J Radiat Oncol Biol Phys 2009;74:1–8.
[19] Tam M, Riaz N, Kannarunimit D, Pena AP, Schupak KD, Gelblum DY, et al. Sparing bilateral neck level IB in oropharyngeal carcinoma and xerostomia outcomes. Am J Clin Oncol 2015;38:343–7.
[20] Eisbruch A, Kim HM, Terrell JE, Marsh LH, Dawson LA, Ship JA. Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. Int J Radiat Oncol Biol Phys 2001;50:695–704.
[21] Robitaille N, Mouiha A, Crépeault B, Valdivia F, Duchesne S. Tissue-based MRI intensity standardization: application to multicentric datasets. Int J Biomed Imaging 2012;2012.
[22] Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015;117:83–90.
[23] Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
[24] Tang X. Texture information in run-length matrices. IEEE Trans Image Process 1998;7:1602–9.
[25] Galloway MM. Texture analysis using gray level run lengths. Comput Graph Image Process 1975;4:172–9.
[26] Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 2015;5:11075.
[27] Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative – feature definitions. Cite as:arXiv:1612.07003 [cs.CV]; 2016.
[28] Dehing-Oberije C, De Ruysscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. Radiother Oncol 2010;97:455–61.
[29] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1.
[30] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing; 2011. Available online at http://www.R-project.org/.
[31] Nardone V, Tini P, Nioche C, Mazzei MA, Carfagno T, Battaglia G, et al. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT. Radiol Medica 2018:1–9.
[32] Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. Front Oncol 2018;8:1–20.
[33] Scalco E, Fiorino C, Cattaneo GM, Sanguineti G, Rizzo G. Texture analysis for the assessment of structural changes in parotid glands induced by radiotherapy. Radiother Oncol 2013;109:384–7.
[34] Broggi S, Fiorino C, Dell'Oca I, Dinapoli N, Paiusco M, Muraglia A, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. Radiother Oncol 2010;94:206–12.
[35] Http://www.startradiology.com/the-basics/mri-technique/. MRI Technique n.d.
[36] Izumi M, Eguchi K, Ohki M, Uetani M, Hayashi K, Kita M, et al. MR imaging of the parotid gland in Sjögren's syndrome: a proposal for new diagnostic criteria. Am J Roentgenol 1996;166.
[37] Izumi M, Hida A, Takagi Y, Kawabe Y, Eguchi K, Nakamura T. MR imaging of the salivary glands in sicca syndrome: comparison of lipid profiles and imaging in patients with hyperlipidemia and patients with Sjogren's syndrome. AJR AmJRoentgenol 2000;175:829–34.
[38] Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomics features on voxel size and number of gray levels Supplementary Material. Med Phys 2017:1–10.
[39] Mackin D, Fave X, Zhang L, Fried D, Taylor B, Rodriguez-rivera E, et al. Measuring CT scanner variability of radiomics features. Invest Radiol 2015;50:757–65.
[40] Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 2000;19:143–50.
[41] Li C, Gore JC, Davatzikos C. Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. Magn Reson Imaging 2014;32:913–23.